



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/703,174	10/31/2000	Charu C. Aggarwal	YOR920000430US1	7445

7590 05/25/2005
William E Lewis
Ryan Mason & Lewis LLP
90 Forest Avenue
Locust Valley, NY 11560

EXAMINER

HILLERY, NATHAN

ART UNIT	PAPER NUMBER
----------	--------------

2176

DATE MAILED: 05/25/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary

Application No.

09/703,174

Applicant(s)

AGGARWAL ET AL.

Examiner

Nathan Hillery

Art Unit

2176

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 30 March 2005.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-27 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-27 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
 - ☐ Certified copies of the priority documents have been received in Application No. _____.
 - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|---|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date. _____ |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

1. This action is responsive to communications: Appeal Brief filed on 3/31/05.
2. Claims 1 – 27 are pending in the case. Claims 1, 10, and 19 are independent.

Response to Arguments

3. In view of the Appeal Brief filed on 3/31/05, PROSECUTION IS HEREBY REOPENED. A new ground of rejection is set forth below.

To avoid abandonment of the application, appellant must exercise one of the following two options:

- (1) file a reply under 37 CFR 1.111 (if this Office action is non-final) or a reply under 37 CFR 1.113 (if this Office action is final); or,
- (2) request reinstatement of the appeal.

If reinstatement of the appeal is requested, such request must be accompanied by a supplemental appeal brief, but no new amendments, affidavits (37 CFR 1.130, 1.131 or 1.132) or other evidence are permitted. See 37 CFR 1.193(b)(2).

4. Applicant's arguments, see pages 5 – 7, filed 3/31/05, with respect to the rejection(s) of claim(s) 1 – 27 under 35 USC 103(a) have been fully considered and are persuasive. Therefore, the rejection has been withdrawn. However, upon further consideration, a new ground(s) of rejection is made in view of a different interpretation of the previously applied reference.
5. In response to applicant's argument that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies (i.e., an intelligent crawling technique that is able to further focus its search

appropriately) are not recited in the rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

6. In response to applicant's argument that Bharat does not disclose an intelligent crawling technique, a recitation of the intended use of the claimed invention must result in a structural difference between the claimed invention and the prior art in order to patentably distinguish the claimed invention from the prior art. If the prior art structure is capable of performing the intended use, then it meets the claim. In a claim drawn to a process of making, the intended use must result in a manipulative difference as compared to the prior art. See *In re Casey*, 152 USPQ 235 (CCPA 1967) and *In re Otto*, 136 USPQ 458, 459 (CCPA 1963).

7. In response to Applicant's arguments that Bharat does not disclose **collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed** nor **using the collected statistical information to automatically determine further document retrieval operations** (pages 5 – 7), it should be noted that upon careful reconsideration, the Office not only maintains that Bharat et al. suggests the limitations but the Office now asserts that Bharat et al. explicitly disclose and teach **collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed** and **using the collected statistical information to automatically determine further document retrieval operations**. Originally, the Office gave the claimed limitations the benefit of the doubt by reading more into the limitations than

were actually claimed. Using the broadest most reasonable interpretation however, the Office notes that “statistical information” simply means *knowledge pertaining to the collection, classification, analysis and interpretation of numerical data* (Source: ISEP / RHW), which can be found at (<http://www.epa.gov/trs/>). Therefore, the equation, $Score(p) = in_degree + 2 * (num_query_matches) + out_degree$, used by Bharat et al. to score pages is statistical information. The equation is used to create a *broader query topic Q 245 in step 240* (Column 6, lines 5 – 6). According to the latter limitation, **using the collected statistical information to automatically determine further document retrieval operations**, the statistical information is simply used to determine further operations; the statistical information is NOT used to perform such operations. Further support that Bharat et al. teach the limitation is evidenced by *when examining a page, we fetch it and compute its relevance, if not previously processed, until five pages have been fetched, or enough top ranked pages have been found relevant, for example, fifteen. In the latter case, the process terminates, and in the former case the process starts a new round until the quota of pages to be fetched is exhausted (step 340), one hundred in our preferred implementation. The last set of rankings determined for hubs and authorities is returned as the result set 112. The motivation for stopping each round when a fixed number of pages, e.g., five in our preferred our implementation, have been fetched is that it is usually sufficient if the top ranked pages are pruned, because these pages tend to be represented by high degree nodes that have a high influence on the ranking of other nodes. After this point, it is more profitable to execute another round than to continue with the pruning* (Column 8, lines 35 – 51).

Claim Rejections - 35 USC § 102

8. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

9. Claims 1 – 27 are rejected under 35 U.S.C. 102(e) as being anticipated by Bharat et al. (US006112203A).

10. **Regarding independent claim 1**, Bharat et al. teach that *in one aspect of the invention, the documents are Web pages connected to each other by hyperlinks. The identities of the documents, and the hyperlinks are in the form of a string called a Uniform Resource Locator (URL). The URLs specify the addresses of the various documents. The set of documents can be produced by combining the set of results from a Web search engine in response to a user query (which we call the 'start-set'), with pages that either link to or are linked from the start-set documents. Terms of the query imply a topic of interest on which the user requested the search to be made. The nodes in the start set are first scored according to their connectivity, and the number of terms of the query that appear as unique sub-strings in the URL of the represented documents. The score is a weighted sum of the number of directed edges to and from a node and the number of unique sub-strings of the URL that match a query term* (Column 2, line 66 – Column 3, line 15), which provide for **retrieving one or more**

documents from the information network that satisfy a user-defined predicate.

Bharat et al. teach that *in our present invention, we use only a subset of the pages for the purpose of content analysis. The subset of influential pages are selected by a heuristic that is based on the URLs of the pages in the start set 201 and their connectivity. This information can be determine from the graph 211 without having to fetch the pages themselves. The heuristic selects nodes based on "in-degree," i.e., the number of edges 213 pointing at a node, "out-degree" (out-going edges) and comparison of the key words in the query with unique sub-strings of the URL. Specifically, in step 220, we score each page p of the input set 201 to determine a value $\text{Score}(p)$ 225. Let n_p be the node representing page p . The score is determined by:*

$$\text{Score}(p) = \text{in_degree} + 2 * (\text{num_query_matches}) + \text{out_degree},$$

*where in_degree is the number of edges pointing at node n_p , num_query_matches is the number of unique sub-strings of the URL of the page p that exactly match a term in the user's query, and out-degree is 1 if the node n_p has at least one edge pointing to another page; otherwise, the value of out-degree is 0. Note, the values $\text{Score}(p)$ 225 can be determined without having to fetch the actual pages (Column 5, lines 47 – 67), which provide for **collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed.** Bharat et al. teach that *in step 230, a small subset of start set pages 235 with the highest values n_p are selected. We select thirty, although it should be understood, that other sized subsets can also be used. The subset of pages 235 is used to distill the broader query topic Q 245 in step 240. Each page of the subset 235 is fetched, and the first, for example, one-thousand words of all of the**

selected pages are concatenated to form Q (Column 6, lines 1 – 8), which provide for using the collected statistical information to automatically determine further document retrieval operations.

11. **Regarding dependent claim 2**, Bharat et al. teach that *in response to a query composed by a user, the search engine returns a result set which satisfies the terms (key words) of the query (Column 4, lines 11 – 14), which provides that the user-defined predicate specifies content associated with a document.*

12. **Regarding dependent claims 3 and 4**, Bharat et al. teach that *during a content analysis phase, a relevance weight is assigned to a carefully chosen subset of the nodes in the graph. The relevance weights are based on the similarity of each represented document to the distilled topic as determined above. The relevance weight of a document is further increased when the document includes words that are terms of the query (Column 3, lines 21 – 27), which provide that the statistical information collection step uses content of the one or more retrieved documents and that the statistical information collection step considers whether the user-defined predicate has been satisfied by the one or more retrieved documents.*

13. **Regarding dependent claims 5 and 6**, Bharat et al. teach that *in step 260, we assign a similarity weight to each node 213 of the sub-graph 255. Various document similarity measuring techniques have been developed in Information Retrieval to determine the goodness of fit between a "target" document and a collection of documents. These techniques typically measure a similarity score based on word frequencies in the collection and a target document (Column 6, lines 51 – 57), which*

provide that **the collected statistical information is used to direct further document retrieval operations toward documents which are similar to the one or more retrieved documents that also satisfy the predicate, and that the collected statistical information is used to direct further document retrieval operations toward documents which are more likely to satisfy the predicate than would otherwise occur with respect to document retrieval operations that are not directed using the collected statistical information.**

14. **Regarding dependent claim 7**, Bharat et al. teach that *in one prior art technique, an algorithm for connectivity analysis of a neighborhood graph (n-graph) is described by Kleinberg ... The algorithm analyzes the link structure, or connectivity of Web pages "in the vicinity" of the result set to suggest useful pages in the context of the search that was performed (Column 1, lines 55 – 64), which provide the capability that* **the collected statistical information is used to direct further document retrieval operations toward documents which are linked to by other documents which also satisfy the predicate.**

15. **Regarding dependent claim 8**, Bharat et al. teach that *FIG. 1 shows a distributed network of computers 100 that can use our invention. Client computers 110 and server computers 120 (hosts) are connected to each other by a network 130, for example, the Internet. The network 130 includes an application level interface called the World Wide Web (the "Web") (Column 3, lines 59 – 64) and that although the invention is described with respect to documents that are Web pages, it should be understood that the invention can also be worked with any linked data objects of a database whose*

content and connectivity can be characterized (Column 4, lines 4 – 8), which provide for
the information network is the World Wide Web and a document is a web page.

16. **Regarding dependent claim 9**, Bharat et al. teach that *in our present invention, we use only a subset of the pages for the purpose of content analysis. The subset of influential pages is selected by a heuristic that is based on the URLs of the pages in the start set 201 and their connectivity. This information can be determined from the graph 211 without having to fetch the pages themselves. The heuristic selects nodes based on "in-degree," i.e., the number of edges 213 pointing at a node, "out-degree" (out-going edges) and comparison of the key words in the query with unique sub-strings of the URL* (Column 5, lines 47 - 56), which provides that **the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages.**

17. **Regarding independent claim 10**, the claim incorporates substantially similar subject matter as claim 1, and is rejected along the same rationale.

18. **Regarding dependent claim 11**, the claim incorporates substantially similar subject matter as claim 2, and is rejected along the same rationale.

19. **Regarding dependent claim 12**, the claim incorporates substantially similar subject matter as claim 3, and is rejected along the same rationale.

20. **Regarding dependent claim 13**, the claim incorporates substantially similar subject matter as claim 4, and is rejected along the same rationale.

21. **Regarding dependent claims 14 and 15**, the claims incorporate substantially similar subject matter as claim 6, and are rejected along the same rationale.


22. **Regarding dependent claim 16**, the claim incorporates substantially similar subject matter as claim 7, and is rejected along the same rationale.
23. **Regarding dependent claim 17**, the claim incorporates substantially similar subject matter as claim 8, and is rejected along the same rationale.
24. **Regarding dependent claim 18**, the claim incorporates substantially similar subject matter as claim 9, and is rejected along the same rationale.
25. **Regarding independent claim 19**, the claim incorporates substantially similar subject matter as claim 1, and is rejected along the same rationale.
26. **Regarding dependent claim 20**, the claim incorporates substantially similar subject matter as claim 2, and is rejected along the same rationale.
27. **Regarding dependent claim 21**, the claim incorporates substantially similar subject matter as claim 3, and is rejected along the same rationale.
28. **Regarding dependent claim 22**, the claim incorporates substantially similar subject matter as claim 4, and is rejected along the same rationale.
29. **Regarding dependent claims 23 and 24**, the claims incorporate substantially similar subject matter as claim 6, and are rejected along the same rationale.
30. **Regarding dependent claim 25**, the claim incorporates substantially similar subject matter as claim 7, and is rejected along the same rationale.
31. **Regarding dependent claim 26**, the claim incorporates substantially similar subject matter as claim 8, and is rejected along the same rationale.
32. **Regarding dependent claim 27**, the claim incorporates substantially similar subject matter as claim 9, and is rejected along the same rationale.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Nathan Hillery whose telephone number is (571) 272-4091. The examiner can normally be reached on M - F, 10:30 a.m. - 7:00 p.m..

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Joseph H. Feild can be reached on (571) 272-4090. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).


JOSEPH FEILD
SUPERVISORY PATENT EXAMINER

NH